

## SUPPLEMENTARY DATA

# DNA backbone interactions impact the sequence specificity of DNA sulfur-binding domains: revelations from structural analyses

**Table S1.** Primers used for construction of protein expression vectors

**Table S2.** Data collection and refinement statistics for SBD<sub>Spr</sub> complexes

**Table S3.** K<sub>D</sub> values of SBD<sub>Spr</sub>-mutant proteins for G<sub>Ps</sub>GCC

**Table S4.** Binding free energy (Units in kcal/mol) between PT-DNA and SBD<sub>Sco</sub>

**Figure S1.** Determination of the binding affinity of SBD<sub>Spr</sub> for R<sub>P</sub> and S<sub>P</sub> stereoisomers of G<sub>Ps</sub>GCC, G<sub>Ps</sub>ATC, G<sub>Ps</sub>AAC, G<sub>Ps</sub>TTC, and C<sub>Ps</sub>CA by fluorescence polarization assay.

**Figure S2.** Superimposition of SBD<sub>Spr</sub> with PT-DNA in different core sequences.

**Figure S3.** Structure of the SBD<sub>Spr</sub> complexed with PT-DNA.

**Figure S4.** Determination of the binding affinity of SBD<sub>Spr</sub>-mutant proteins for G<sub>Ps</sub>GCC through EMSA and fluorescence polarization assay.

**Figure S5.** Superimposition of the SBD<sub>Spr</sub>-DNA complex (in cyan and orange) and SBD<sub>Sco</sub>-DNA complex (in grey and yellow).

**Figure S6.** Superimposition of the hydrophobic cavity in SBD<sub>Spr</sub> and SBD<sub>Sco</sub>.

**Figure S7.** The ligand-receptor interaction of methyl groups in the T6=A6' and A6=T6' base pairs in G<sub>Ps</sub>ATC and G<sub>Ps</sub>AAC.

**Figure S8.** Determination of the binding affinity of SBD<sub>Sco</sub> mutants.

**Figure S9.** Noncovalent interaction (NCI) analysis of the interface domain in SBD<sub>Sco</sub>-G<sub>PS</sub>GCC, SBD<sub>Sco</sub>-G<sub>PS</sub>ATC and SBD<sub>Sco</sub>-G<sub>PS</sub>AAC.

**Figure S10.** Structural comparison of representative structures of SBD<sub>Sco</sub>-E156R/D157R and wild type extracted from clustering analysis.

**Figure S11.** Ability of the SBD<sub>Sco</sub>-E156R/D157R mutant to bind *R<sub>P</sub>* and *S<sub>P</sub>* stereoisomers of PT-DNA.

**Figure S12.** Multiple sequence alignment of SBD homologs.

**Table S1.** Primers and oligonucleotides used in this study

PRIMERS	SEQUENCE
SBDspr-F	GGACCATATGCCGCTCACCGACACAGACCGGT
SBDspr-F	GGAATTCTTAGCCTCGTCGTACAGGCCGG
SBDspr-Y31A-F	CGCGCCTC GCC CAGCCGATCACTCTGCT
SBDspr-Y31A-R	ATCGGCTGGCGAGGCAGCGGTCCGTTCG
SBDspr-Q32A-F	GCCTCTAT GCG CCGATCACTCTGCTGTG
SBDspr-Q32A-R	GTGATCGGCGCATAGAGGCGCGGTCCGT
SBDspr-Y78A-F	GCCCCGGAC GCC CCCGTCCTCGCGCTCCA
SBDspr-Y78A-R	AGGACGGGGCGTCCGGCGGGGGCGCT
SBDspr-P79A-F	CGGACTAC GCC GTCCTCGCGCTCCACCG
SBDspr-P79A-R	GCGAGGACGGCGTAGTCCGGCGGGGC
SBDspr-A82G-F	CCGTCCTC GGG CTCCACCGCGCGGGGTT
SBDspr-A82G-R	CGGTGGAGCCCCGAGGACGGGGTAGTCCG
SBDspr-H102A-F	CCACCGCG GCC GGCGACTCGCGCTCAG
SBDspr-H102A-R	GAGTCGCCGCCGCGGTGGCACCTCGC
SBDspr-G103A-F	CCGCGCAC GCC GACTCGCGCTCAGGAA
SBDspr-G103A-R	GCCGAGTCGGCGTGCACGGTGACCGT
SBDspr-D104A-F	CGCACGGC GCC TCGCGCTCAGGAACGT
SBDspr-D104A-R	AGCGCCGAGGCGCCGTGCGCGGTGGCA
SBDspr-R73A-F	GGGGGGAG GCC CCCCAGCCGGACTACCC
SBDspr-R73A-R	GGGCGGGGGGCCTCCCCCCCAGGCCCCGT
SBDspr-R85A-F	CGCTCCAC GCC GCAGGGTTGTGGACGCT
SBDspr-R85A-R	AACCCCGGGCGTGGAGCGCGAGGACGG
SBDsco-F	GGAC CATATG ATCAGGGAGGCCCCAAGACCT
SBDsco-R	GGAATTCTTAGTGGTGGTGGTGGTGCAGAGCA TCCGGCCA
SBDsco-E156R-F	GCCAGGTT CGA GACGGTGTGGATGGGGT
SBDsco-E156R-R	ACACCGTCTCGAACCTGGCGTACTTCT
SBDsco-E156K-F	GCCAGGTT AAG GACGGTGTGGATGGGGT
SBDsco-E156K-R	ACACCGTC CTT AACCTGGCGTACTTCT
SBDsco-E156L-F	GCCAGGTT TTA GACGGTGTGGATGGGGT
SBDsco-E156L-R	ACACCGTC TAA AACCTGGCGTACTTCT
SBDsco-E156Q-F	GCCAGGTT CAA GACGGTGTGGATGGGGT
SBDsco-E156Q-R	ACACCGTC TTG AACCTGGCGTACTTCT
SBDsco-E156D-F	GCCAGGTT GAT GACGGTGTGGATGGGGT
SBDsco-E156D-R	ACACCGTC ATC AACCTGGCGTACTTCT
SBDsco-D157R-F	CAGGTTGAAAGGGGTGTGGATGGGTGCG
SBDsco-D157R-R	ATCCACACCCCTTCAACCTGGCGTACT
SBDsco-D160R-F	GTTGAAGACGGTGTGAGAGGGGTGCGCTACCCATTCTGGGC
SBDsco-D160R-R	TGGGTAGCGCACCCCTCTCACACCGTCTAACCTGGCGT
SBDsco-E156R/D157R-F	CAGGTTCGAAGGGGTGTGGATGGGTGCG
SBDsco-E156R/D157R-R	ATCCACACCCCTTCAACCTGGCGTACT

---

<b>OLIGONUCLEOTIDES</b>	<b>SEQUENCE</b>
PT-GGCC-8	5'-GGCG <sub>Ps</sub> GCCC-3' 3'-CCGCCGGG-5'
PT-GGCC-10	5'-CCCG <sub>Ps</sub> GCCGCC-3' 3'-GGGCCGGCGG-5'
GGCC-10	5'-CCCGGCCGCC-3' 3'-GGGCCGGCGG-5'
PT-GATC-8	5'-GATG <sub>Ps</sub> ATCC-3' 3'-CTACTAGG-5'
PT-GATC-10	5'-GATG <sub>Ps</sub> ATCCTA-3' 3'-CTACTAGGAT-5'
GATC-10	5'-GATGATCCTA-3' 3'-CTACTAGGAT-5'
PT-GAAC-10	5'-GGCG <sub>Ps</sub> AACGTG-3' 3'-CCGCTTGCAC-5'
GAAC-10	5'-GGCGAACGTG-3' 3'-CCGCTTGCAC-5'
PT-GTTC-10	5'-GGCGAACGTG-3' 3'-CCGCTT <sub>Ps</sub> GCAC-5'
PT-CCA-10	5'-GAACC <sub>Ps</sub> CAAGC-3' 3'-CTTGGGTTCG-5'
CCA-10	5'-GAACCCAAGC-3' 3'-CTTGGGTTCG-5'

---

**Table S2.** Data collection and refinement statistics for SBD<sub>Spr</sub> complexes

	<b>SBD<sub>Spr</sub>-GPSGCC complex</b>	<b>SBD<sub>Spr</sub>-GPSATC complex</b>	<b>SBD<sub>Spr</sub>-GPSAAC complex</b>
<b>Data collection</b>			
Beamline	BL19U1	BL19U1	BL18U1
Wavelength (Å)	0.97925	0.97925	0.97776
Resolution range*	48.13 - 2.063 (2.137 - 2.063)	31.12 - 3.3 (3.417 - 3.3)	28.63 - 2.42 (2.506 - 2.42)
Space group	<i>P</i> 6 <sub>1</sub> 22	<i>C</i> 222 <sub>1</sub>	<i>P</i> 1
Unit cell parameters			
a, b, c (Å)	154.5, 154.5, 156.8	97.7, 105.3, 116.4	46.94, 48.94, 56.19
α, β, γ (°)	90, 90, 120	90, 90, 90	107.04, 109.79, 97.26
Unique reflections*	68202 (6703)	9259 (922)	13319 (1393)
Completeness (%)*	100(100)	100(100)	80 (85)
<i>I</i> / <i>σI</i> *	32.95 (1.53)	6.5 (1.5)	80.29 (1.47)
Redundancy*	38.3(29.8)	13.1 (13.9)	3.5 (3.5)
R-merge (%)*	14.2 (247.7)	45.0 (155.1)	7.4 (52.7)
<b>Refinement</b>			
Resolution range	48.13 - 2.063	31.12 - 3.3	28.63 - 2.42
Average B-factor	36.12	62.13	47.80
R <sub>work</sub> /R <sub>free</sub> factors (%)	20.98/23.04	24.47/29.54	19.99/24.77
RMSD bond lengths (Å)	0.005	0.006	0.007
RMSD bond angles (°)	0.75	0.75	0.97
Ramachandran plot (favoured, allowed, outliers, %)	98, 1.9, 0	93, 6.3, 0.32	98, 1.8, 0.61
PDB code	7CC9	7CCJ	7CCD

\*Statistics for the highest-resolution shell are shown in parentheses.

R-merge =  $\sum |I - \langle I \rangle| / \sum I$ , where *I* is the observed intensity and  $\langle I \rangle$  is the averaged intensity from multiple observations.

$\langle I/\sigma I \rangle$  = averaged ratio of the intensity (*I*) to the error of the intensity ( $\sigma I$ ).

R<sub>work</sub> =  $\sum |F_{\text{obs}} - F_{\text{cal}}| / \sum |F_{\text{obs}}|$ , where *F*<sub>obs</sub> and *F*<sub>cal</sub> are the observed and calculated structure

factors, respectively.

$R_{\text{free}}$  was calculated using a randomly chosen subset (5%) of the reflections not used in refinement.

RMSD, root-mean-square deviation from ideal geometry. Data for the highest resolution shell are shown in parentheses.

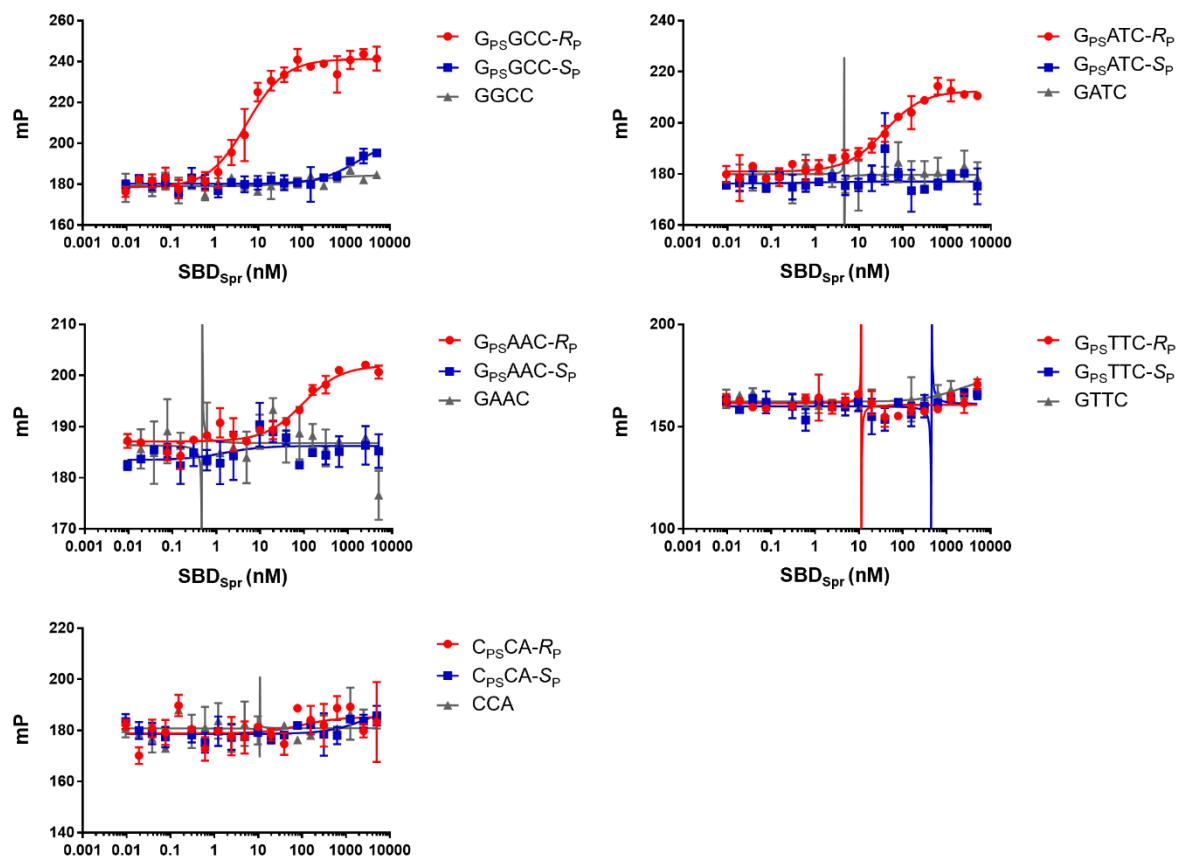
Abbreviations: PDB, Protein Data Bank; RMSD, root mean square deviation.

**Table S3.** K<sub>D</sub> values of SBD<sub>Spr</sub>-mutant proteins for GpsGCC

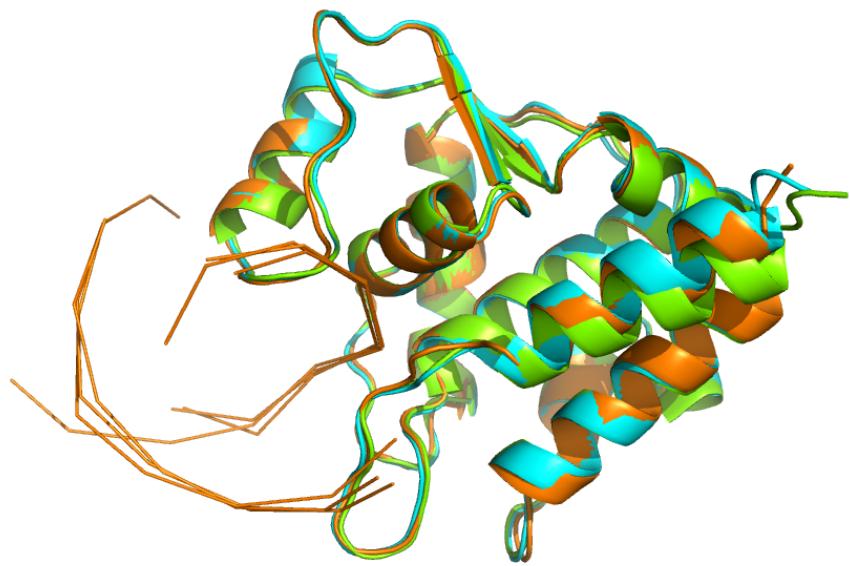
<b>SBD<sub>Spr</sub> mutant</b>	<b>K<sub>D</sub> (nM)</b>
(wild type)	5.6±0.9
Y31A	>3156
Q32A	279±49
Y78A	>4475
P79A	67±8
H102A	>1857
G103A	>9000
D104A	144±12
R73A	>1252
R85A	31±2

**Table S4.** Binding free energy (Units in kcal/mol) between PT-DNA and SBD<sub>Sco</sub>

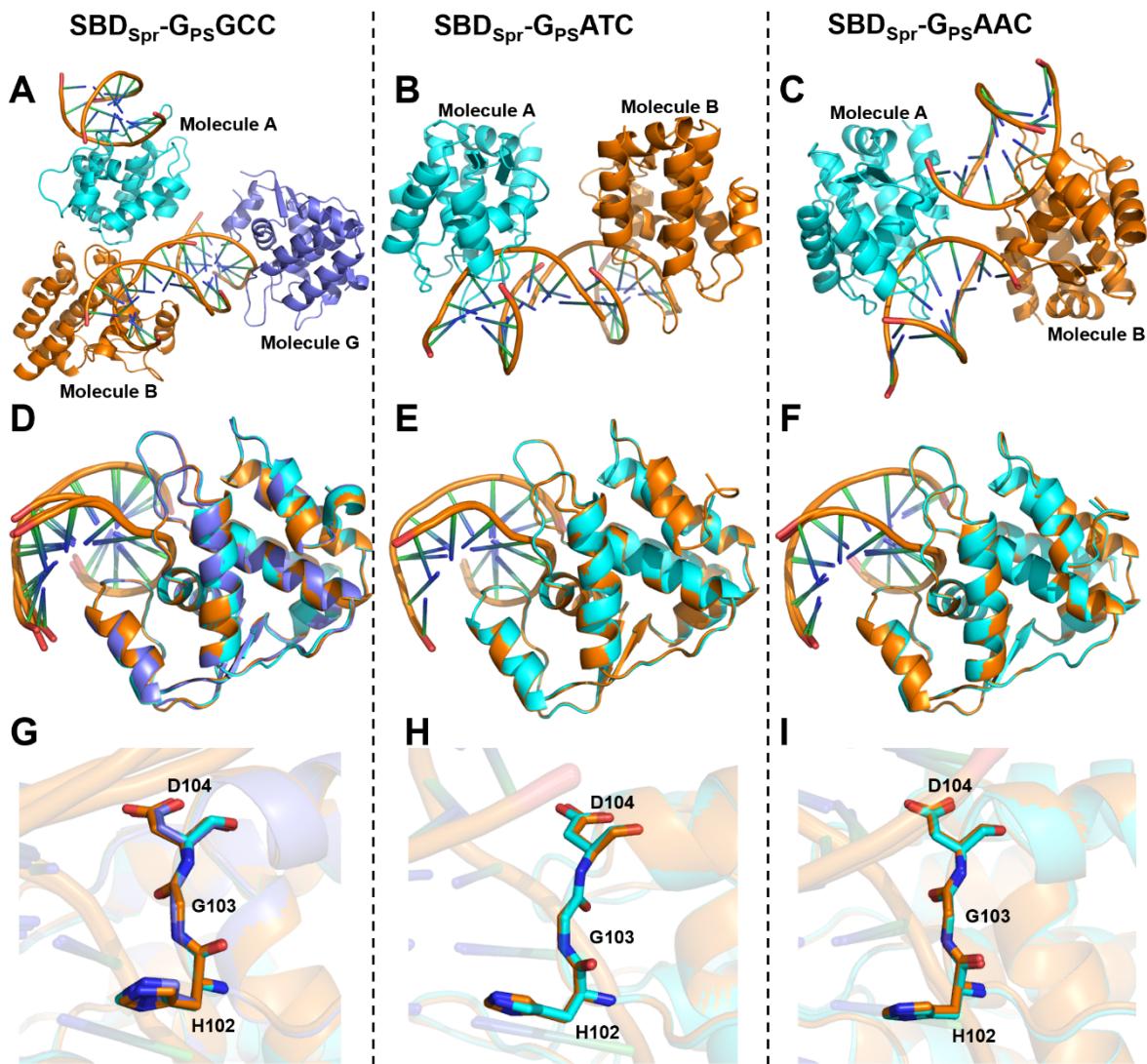
System	$\Delta E_{\text{vdW}}$	$\Delta E_{\text{ele}}$	$\Delta G_{\text{GB}}$	$\Delta G_{\text{SA}}$	$\Delta G_{\text{bind}}$	$\Delta \Delta G_{\text{bind}}$
SBD <sub>Sco</sub>	-98.6 ± 7.4	-1260.9 ± 79.8	1288.4 ± 76.6	-13.3 ± 0.8	-84.4 ± 4.4	0
SBD <sub>Sco-</sub> E156R/D157R	-76.2 ± 7.5	-1207.7 ± 79.8	1212.5 ± 76.6	-11.1 ± 0.8	-83.5 ± 4.8	0.9 ± 0.4



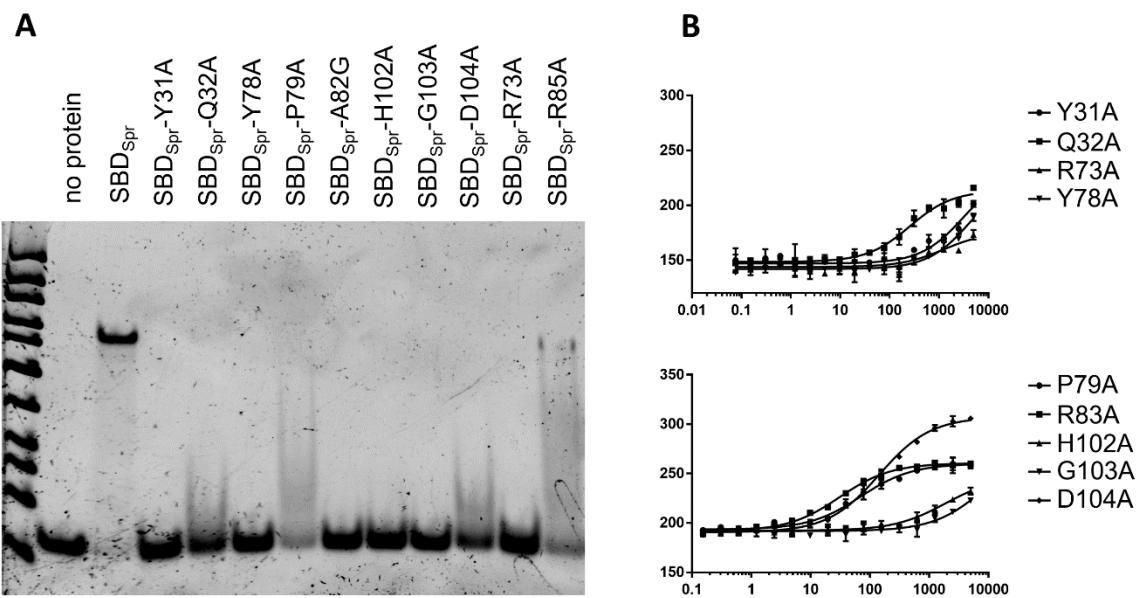
**Figure S1.** Determination of the binding affinity of SBD<sub>Spr</sub> for R<sub>P</sub> and S<sub>P</sub> stereoisomers of G<sub>PS</sub>GCC, G<sub>PS</sub>ATC, G<sub>PS</sub>AAC, G<sub>PS</sub>TTC, and C<sub>PS</sub>CA by fluorescence polarization assay.



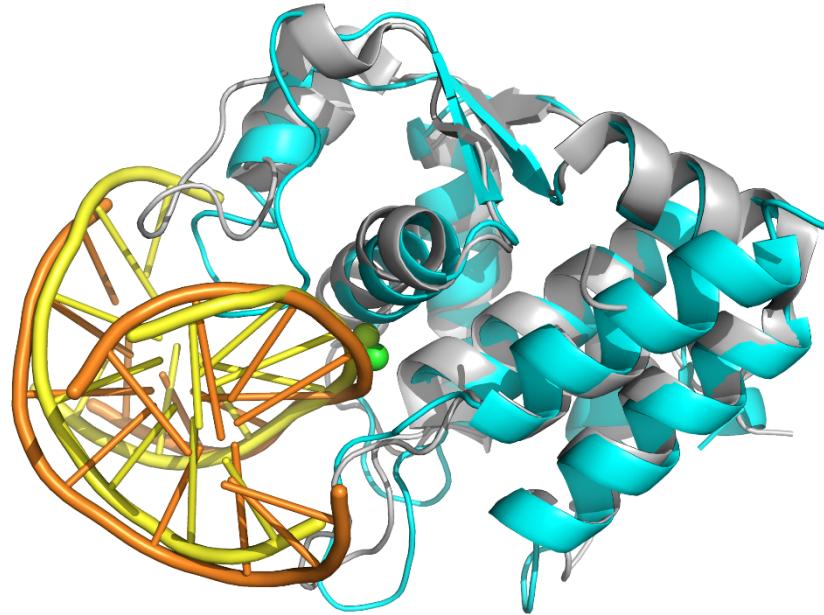
**Figure S2.** Superimposition of SBD<sub>Spr</sub> with PT-DNA of different core sequences. SBD<sub>Spr</sub>-G<sub>Ps</sub>GCC, SBD<sub>Spr</sub>-G<sub>Ps</sub>ATC, and SBD<sub>Spr</sub>-G<sub>Ps</sub>AAC are colored in cyan, orange, and green, respectively.



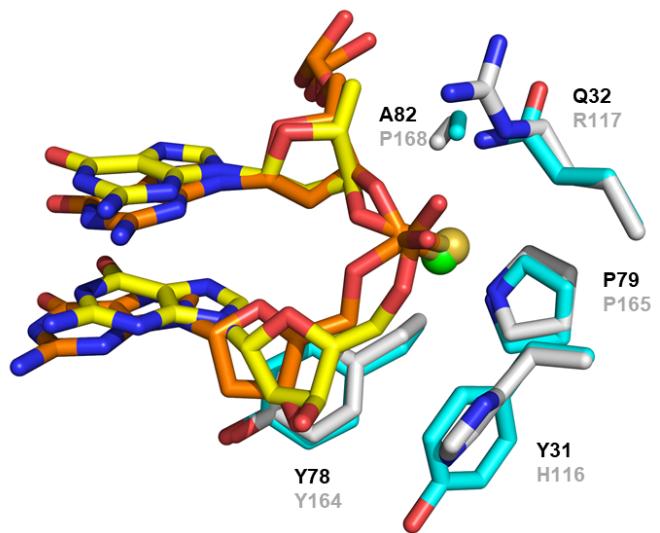
**Figure S3.** Structures of the SBD<sub>Spr</sub> complexed with PT-DNAs. **(A)** Trimer form of the SBD<sub>Spr</sub>-G<sub>Ps</sub>GCC complexes. **(B)** Dimer form of the SBD<sub>Spr</sub>-G<sub>Ps</sub>ATC complexes. **(C)** Dimer form of the SBD<sub>Spr</sub>-G<sub>Ps</sub>ACC complexes. **(D)** Superimposition of molecule A, B, and G of SBD<sub>Spr</sub>-G<sub>Ps</sub>GCC complexes, yielding an RMSD of 0.140 Å. **(E)** Superimposition of molecule A and B of SBD<sub>Spr</sub>-G<sub>Ps</sub>ATC complexes, yielding an RMSD of 0.167 Å. **(F)** Superimposition of molecule A and B of SBD<sub>Spr</sub>-G<sub>Ps</sub>AAC complexes, yielding an RMSD of 0.250 Å. **(G, H, I)** Superimposition of HGD loop in different molecule of SBD<sub>Spr</sub>-G<sub>Ps</sub>GCC complexes, SBD<sub>Spr</sub>-G<sub>Ps</sub>ATC complexes and SBD<sub>Spr</sub>-G<sub>Ps</sub>AAC complexes, respectively.



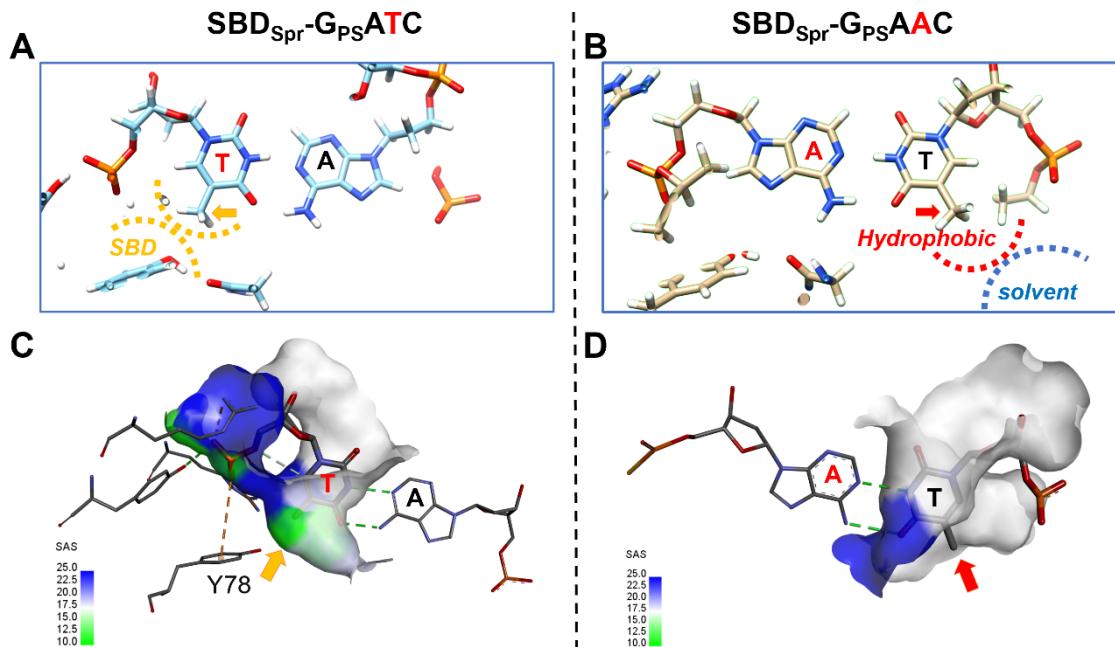
**Figure S4.** Determination of the binding affinity of SBD<sub>Spr</sub> mutant proteins for GPSGCC through **(A)** EMSA and **(B)** fluorescence polarization assay.



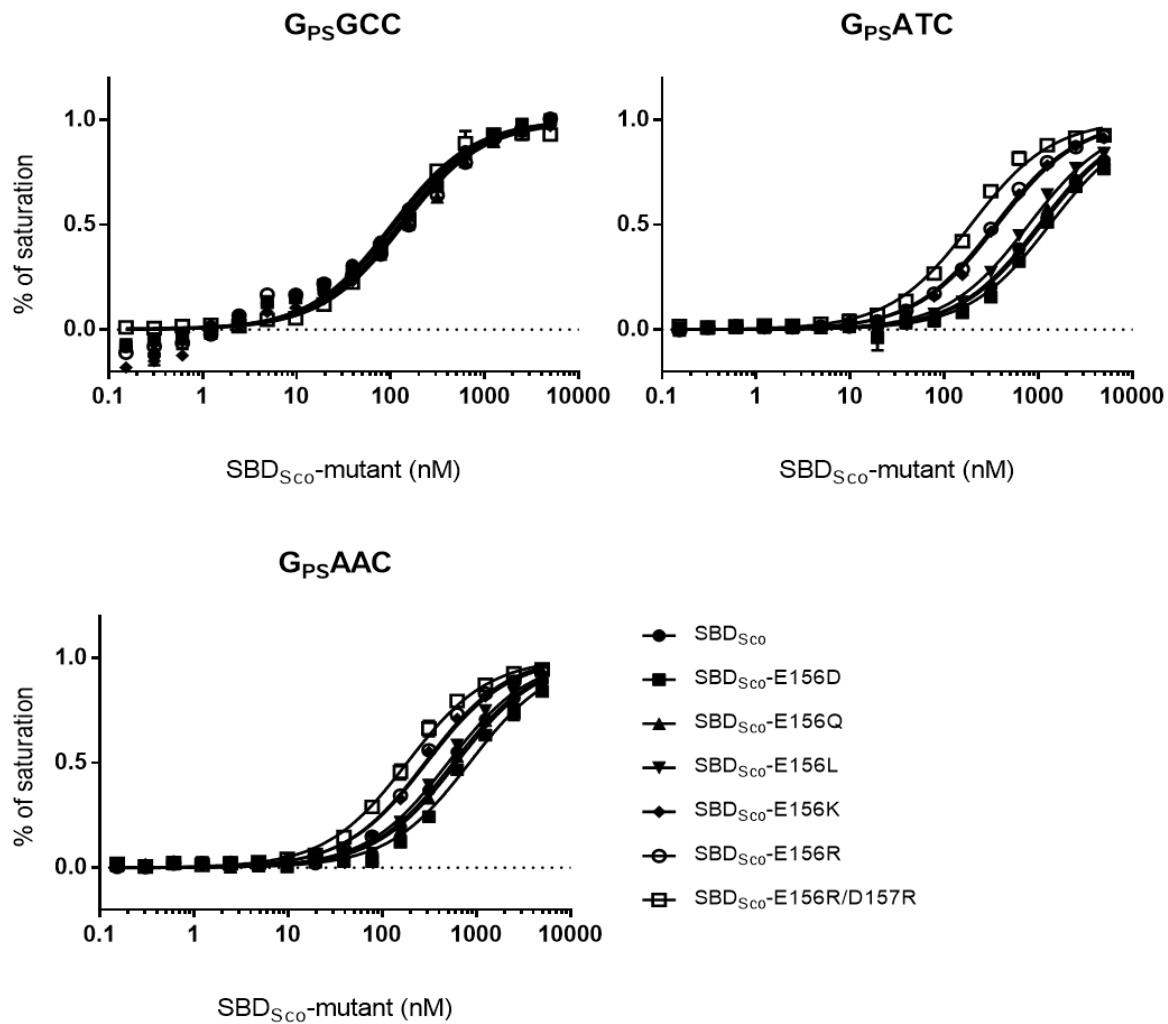
**Figure S5.** Superimposition of the SBD<sub>Spr</sub>-DNA complex (in cyan and orange) and SBD<sub>Sco</sub>-DNA complex (in grey and yellow).



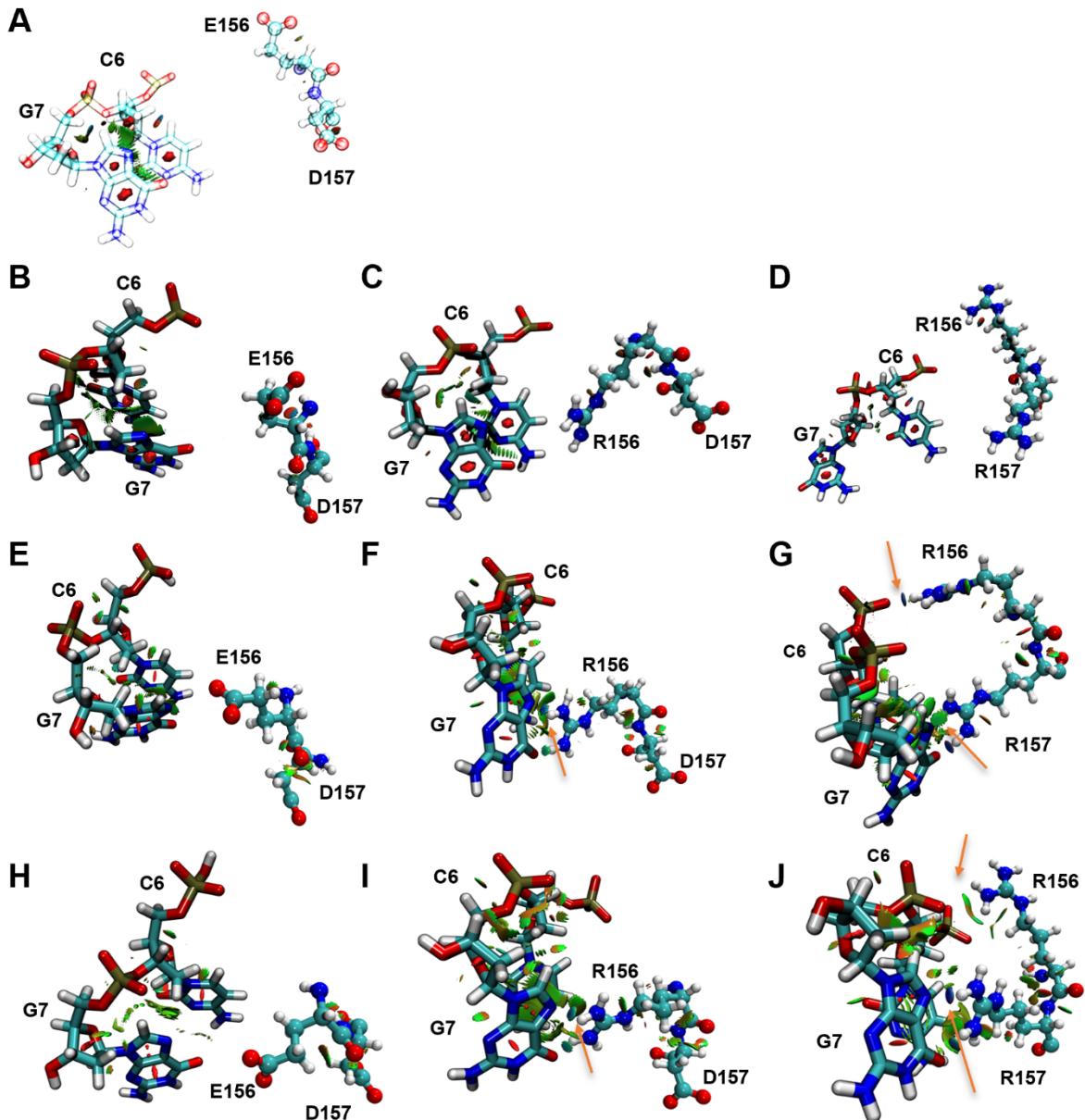
**Figure S6.** Superimposition of the hydrophobic cavity in SBD<sub>sPr</sub> (cyan) and SBD<sub>sCo</sub> (grey).



**Figure S7.** The ligand-receptor interaction of methyl groups in the T6=A6' and A6=T6' base pairs in G<sub>Ps</sub>ATC and G<sub>Ps</sub>AAC. (A, B) Schematic diagram of the hydrophobic interaction in SBD<sub>Spr</sub>-G<sub>Ps</sub>ATC and SBD<sub>Spr</sub>-G<sub>Ps</sub>AAC complex. (C, D) The solvent accessibility surface (SAS) of T<sup>6</sup> in SBD<sub>Spr</sub>-G<sub>Ps</sub>ATC and T<sup>6'</sup> SBD<sub>Spr</sub>-G<sub>Ps</sub>AAC complex, using the percent solvent accessibility in a scale of 10-25 for a better color contrast. Surface was colored by the solvent accessibility of the receptor residues from blue for exposed to green for buried. The arrows pointed to the methyl group of thymine.

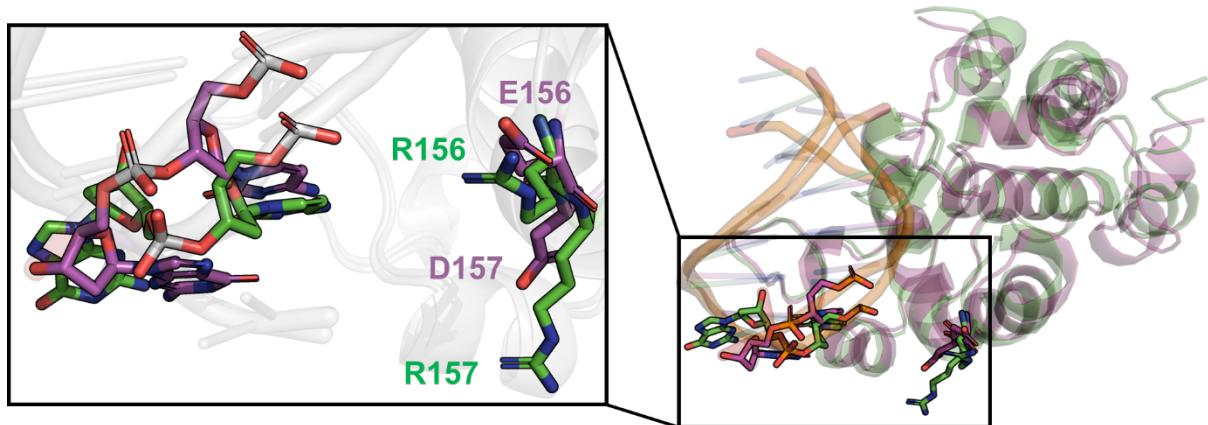


**Figure S8.** Determination of the binding affinity of SBD<sub>Sco</sub> mutants by fluorescence polarization assay. Mutations were generated at sites E156 and D157 in SBD<sub>Sco</sub>, and the binding affinities of the resulting mutants for G<sub>PS</sub>GCC, G<sub>PS</sub>ATC, and G<sub>PS</sub>AAC were calculated and are listed in Table 2.

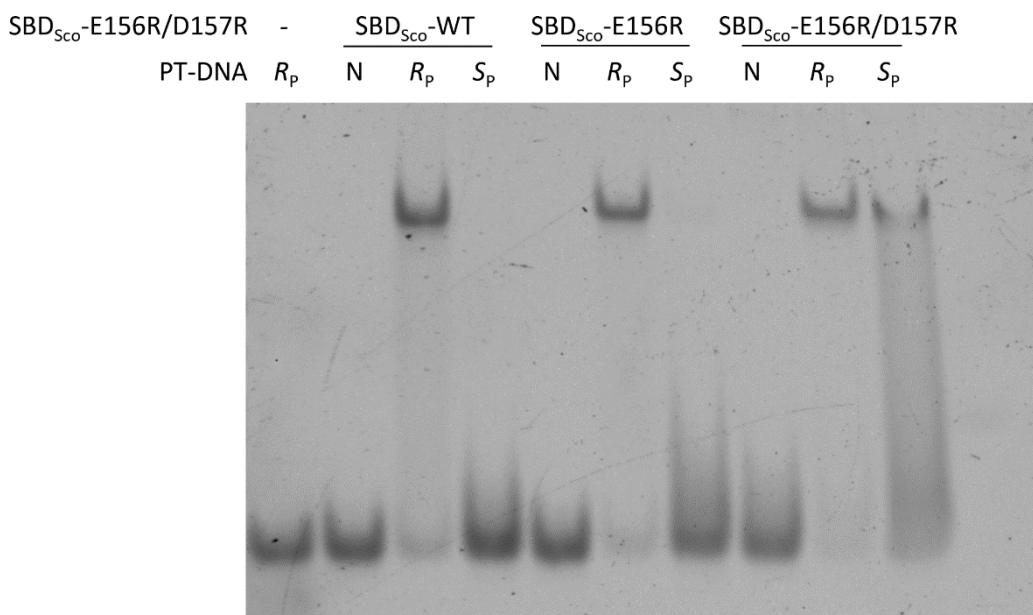


**Figure S9.** Noncovalent interaction (NCI) analysis of the interface domain in SBD<sub>Sco</sub>-G<sub>PS</sub>GCC, SBD<sub>Sco</sub>-G<sub>PS</sub>ATC and SBD<sub>Sco</sub>-G<sub>PS</sub>AAC using reduced density gradient (RDG) method. (A) SBD<sub>Sco</sub>-G<sub>PS</sub>GCC crystal structure, showed as transparent. Panel (B) wild-type SBD<sub>Sco</sub>; (C) SBD<sub>Sco</sub>-E156R mutant; (D) SBD<sub>Sco</sub>-E156R/D157R mutant in complex with G<sub>PS</sub>GCC after MD simulations. Panel (E) wild-type SBD<sub>Sco</sub>; (F) SBD<sub>Sco</sub>-E156R mutant; (G) SBD<sub>Sco</sub>-E156R/D157R mutant in complex with G<sub>PS</sub>ATC after MD simulations. Panel (H) wild-type SBD<sub>Sco</sub>; (I) SBD<sub>Sco</sub>-E156R mutant; (J) SBD<sub>Sco</sub>-E156R/D157R mutant in complex with G<sub>PS</sub>AAC after MD simulations.

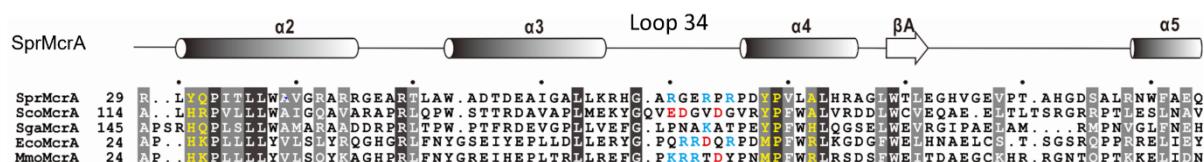
GpsAAC after MD simulations. Plots are given with an isovalue of 0.05 a.u. Arrows indicate the interaction regions between nucleotides and amino acid residues. RDG isosurfaces were colored according to the value of  $\text{sign}(\lambda_2) \rho$ , where a BGR (blue-green-red) color scale was adopted. Blue color represents attractive or bonding interaction, green weak van der Waals interaction, and red repulsive interaction. All isosurfaces are colored according to a BGR scheme over the electron density range  $-0.05 \text{ a.u.} < \text{sign}(\lambda_2) \rho < 0.05 \text{ a.u.}$



**Figure S10.** Structural comparison of representative structures of SBD<sub>Sco</sub>-E156R/D157R and wild type extracted from clustering analysis. Superposition of SBD<sub>Sco</sub>-E156R/D157R structure (green) with wild type (purple).



**Figure S11.** Ability of the SBD<sub>Sco</sub>-E156R/D157R mutant to bind *R<sub>P</sub>* and *S<sub>P</sub>* stereoisomers of hemi-PT-DNA. A 10 bp PT-DNA oligonucleotide (5'-CCCG<sub>Ps</sub>GCCGCC-3') was used as the DNA substrate in the EMSA. N, DNA oligonucleotide with no modification; *R<sub>P</sub>*, *R<sub>P</sub>* stereoisomers of hemi-PT-DNA; *S<sub>P</sub>*, *S<sub>P</sub>* stereoisomers of hemi-PT-DNA.



**Figure S12.** Multiple sequence alignment of SBD homologs. SprMcrA from *Streptomyces pristinaespiralis* was aligned with homologs from *Streptomyces coelicolor* (ScoMcrA), *Streptomyces gancidicus* (SgaMcrA), *Escherichia coli* (EcoMcrA), and *Morganella morganii* (MmoMcrA). The basic amino acid residues of loop 34 are highlighted in blue, and acidic amino acid residues are highlighted in red. The sulfur-recognizing residues are colored in yellow. Shading indicates conserved residues. The secondary structural features of SprMcrA are shown above the alignment.